澳 門 大 學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

認知與腦科學研究中心
Centro de Ciências Cognitivas e Cerebrais
Centre for Cognitive and Brain Sciences

社 會 科 學 學 院
FACULDADE DE CIÊNCIAS SOCIAIS
FACULTY OF SOCIAL SCIENCES

Towards human-compatible autonomous car: A study of modified Turing test in automated driving with affective variability modelling

Presenter: Zhaoning Li 李肇宁

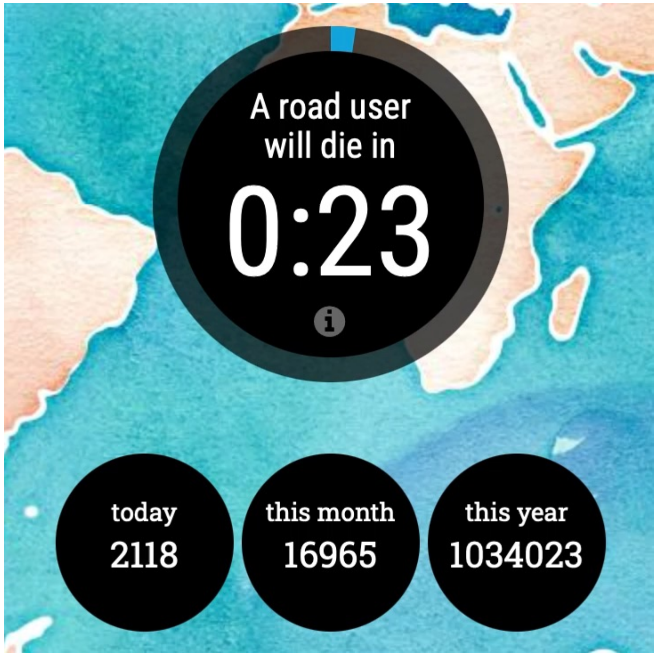
Authors: Zhaoning Li[1,2], Qiaoli Jiang[1], Zhengming Wu[3], Anqi Liu[4], Haiyan Wu[2], Miner Huang[1], Kai Huang[5] and Yixuan Ku[1]

[1]Centre for Brain and Mental Well-being, Department of Psychology, Sun Yat-sen University,
[2]Centre for Cognitive and Brain Sciences and Department of Psychology, University of Macau,
[3]Guangzhou Intelligent Connected Vehicle Pilot Zone Operations Centre,
[4]Department of Computer Science, Whiting School of Engineering, Johns Hopkins University,
[5]School of Computer Science and Engineering, Sun Yat-Sen University

INTERNATIONAL GRADUATE FORUM ON LANGUAGE COGNITIVE SCIENCE

Background

1, 350, 000*



A road user will die in

0:23

today 2118

this month 16965

this year 1034023

Automated driving have the potential to increase road safety, as they can react faster than human drivers and are not subject to human errors.

* World Health Organization. (2018). Global status report on road safety 2018.

Background

Despite the potential benefits, there is no large scale deployment of autonomous cars (ACs) yet.

Existing literature has highlighted that the acceptance of the AC will increase if it drives in a human-like manner.

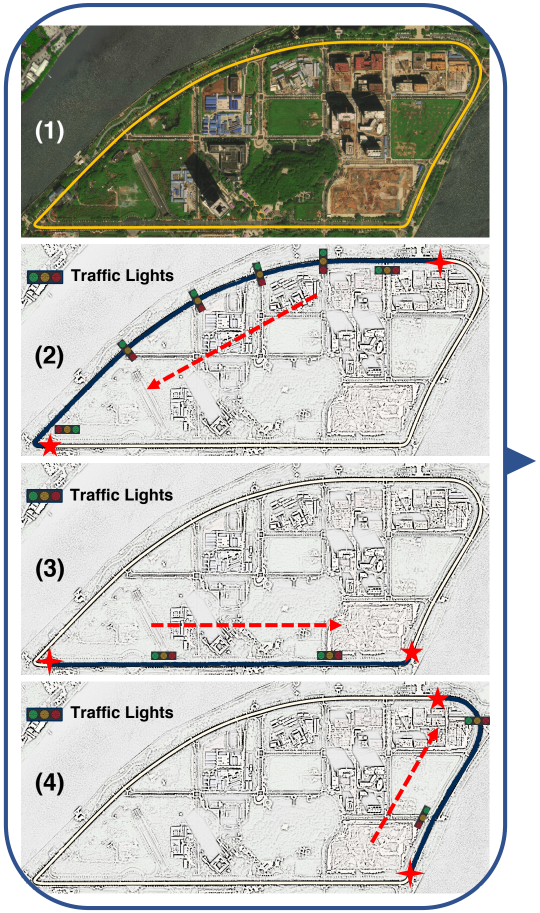
However, literature presents no human-subject research focusing on passengers in a natural environment that examines whether the AC should behave in a human-like manner.

How to offer naturalistic experiences from a passenger's seat perspective to measure the people's acceptance of ACs?

The modified Turing test of automated driving

A

B



C

Results of the modified Turing test

Confusion matrix of three road stages for the results in the Turing test

		Human driver	AI driver	Human driver	AI driver	Human driver	AI driver
Unlikely	***1***	6	8	6	10	11	6
Somewhat likely	***2***	15	9	4	14	13	6
Very likely	***3***	10	20	10	24	9	20
(to be driven by the AI driver)		**First stage 38.24%**		**Second stage 44.12%**		**Third stage 47.69%**	

How do human passengers choose in the modified Turing test of automated driving?

How do human passengers choose?

Passenger

Human driver

or

AI driver

First-order mentalising
(self-other mentalising)

(Wu et al, 2019; Wu et al, 2020)

Choice behaviour →

$$B = f(P, E)$$

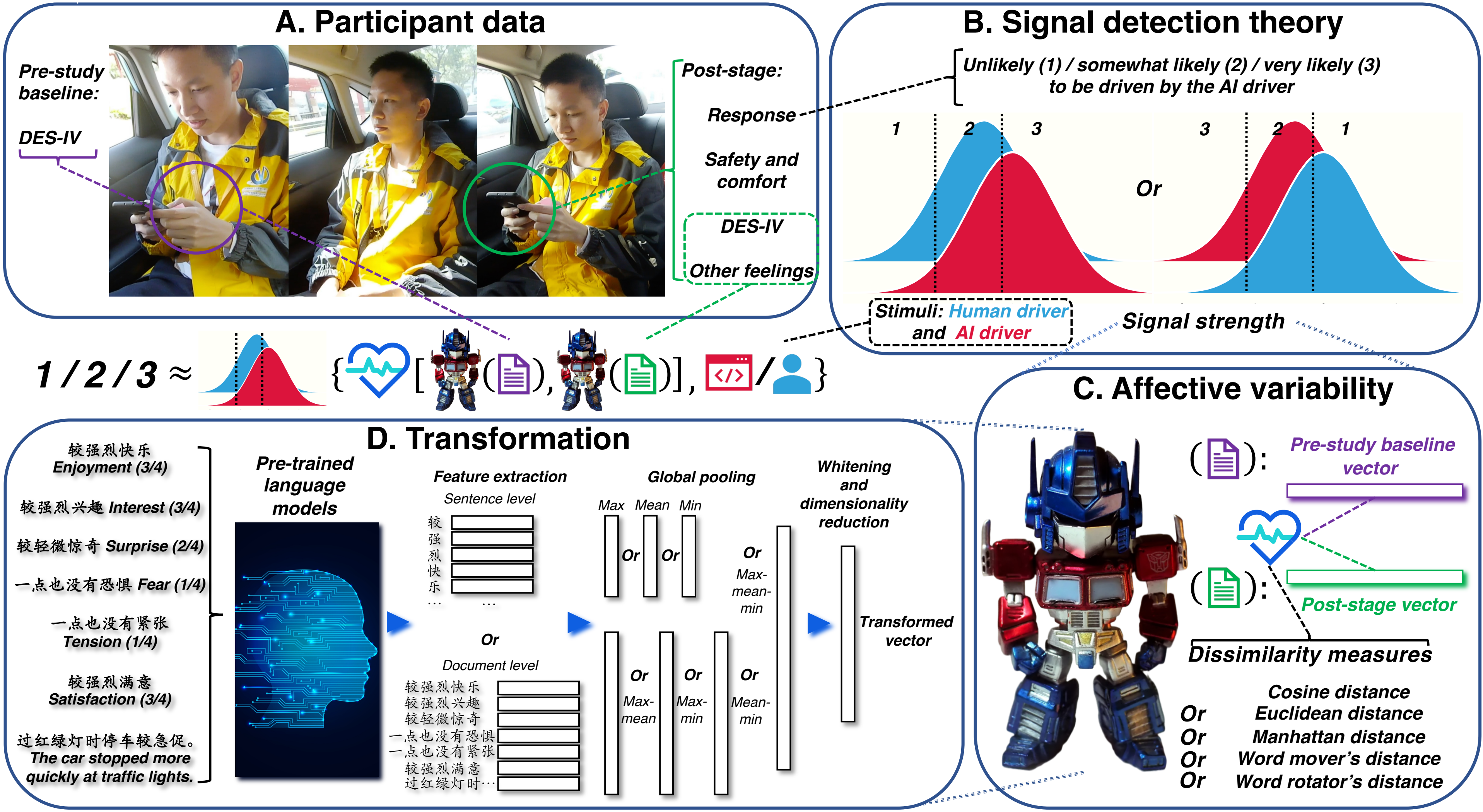
Passenger

Driving environment

Kurt Lewin, (1936)

(Adapted from Wikipedia)

How do human passengers choose?



A. Participant data

Pre-study baseline:

DES-IV

Post-stage:

Response

Safety and comfort

DES-IV

Other feelings

B. Signal detection theory

Unlikely (1) / somewhat likely (2) / very likely (3) to be driven by the AI driver

Stimuli: Human driver and AI driver

Signal strength

1 / 2 / 3 ≈ {♡[(📄), (📄)], / 👤}

C. Affective variability

(📄): *Pre-study baseline vector*

(📄): *Post-stage vector*

Dissimilarity measures

Cosine distance
Or Euclidean distance
Or Manhattan distance
Or Word mover's distance
Or Word rotator's distance

D. Transformation

较强烈快乐
Enjoyment (3/4)

较强烈兴趣 *Interest (3/4)*

较轻微惊奇 *Surprise (2/4)*

一点也没有恐惧 *Fear (1/4)*

一点也没有紧张
Tension (1/4)

较强烈满意
Satisfaction (3/4)

过红绿灯时停车较急促。
The car stopped more quickly at traffic lights.

Pre-trained language models

Feature extraction

Sentence level

Or

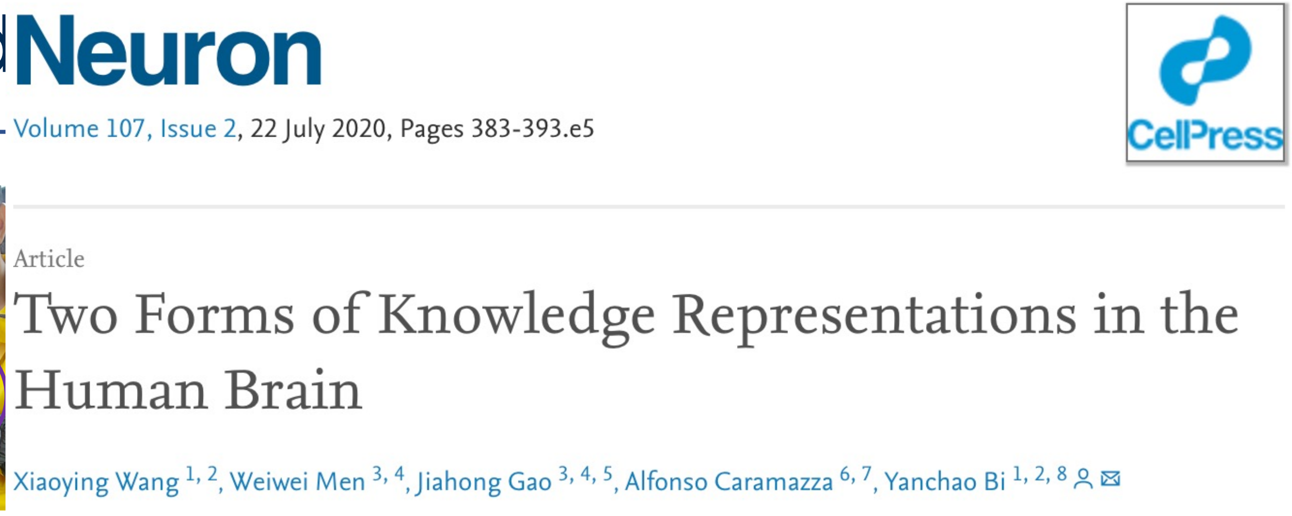
Document level

Global pooling

Whitening and dimensionality reduction

Transformed vector

How d



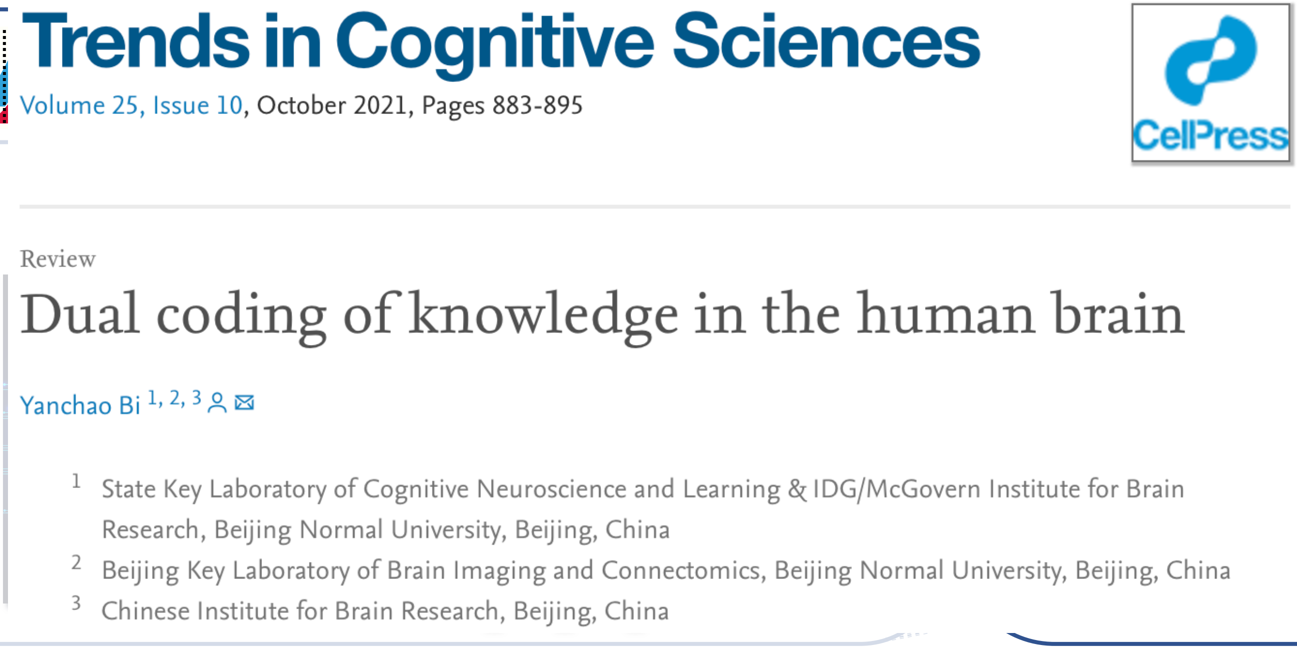
Neuron

Volume 107, Issue 2, 22 July 2020, Pages 383-393.e5

Article

Two Forms of Knowledge Representations in the Human Brain

Xiaoying Wang [1, 2], Weiwei Men [3, 4], Jiahong Gao [3, 4, 5], Alfonso Caramazza [6, 7], Yanchao Bi [1, 2, 8]



Trends in Cognitive Sciences

Volume 25, Issue 10, October 2021, Pages 883-895

Review

Dual coding of knowledge in the human brain

Yanchao Bi [1, 2, 3]

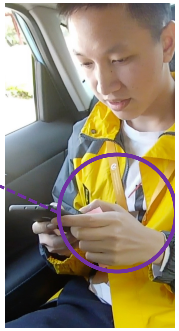
[1] State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China

[2] Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing, China

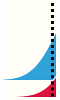
[3] Chinese Institute for Brain Research, Beijing, China

Pre-study baseline:

DES-IV



1 / 2 / 3 ≈



较强烈快乐 *Enjoyment (3/4)*

较强烈兴趣 *Interest (3/4)*

较轻微惊奇 *Surprise (2/4)*

一点也没有恐惧 *Fear (1/4)*

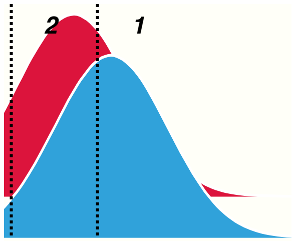
一点也没有紧张 *Tension (1/4)*

较强烈满意 *Satisfaction (3/4)*

过红绿灯时停车较急促。*The car stopped more quickly at traffic lights.*

theory

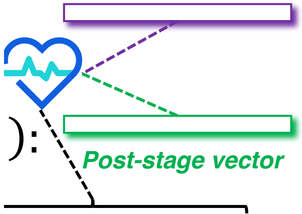
) / very likely (3)
driver



ngth

variability

): *Pre-study baseline vector*



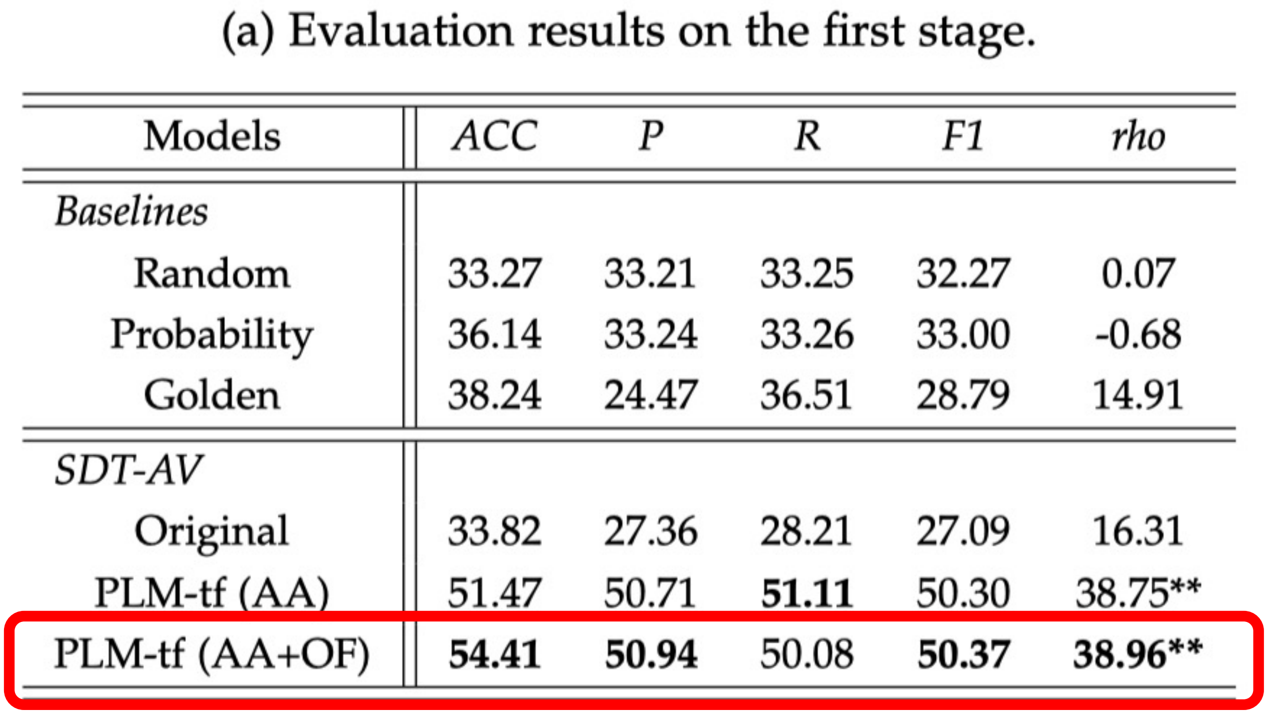
): *Post-stage vector*

milarity measures

Cosine distance
Euclidean distance
Manhattan distance
Word mover's distance
Word rotator's distance

Results of the computational models

Comparison on the Outer Loop Cross-Validation of Nested-LOOCV with Baselines

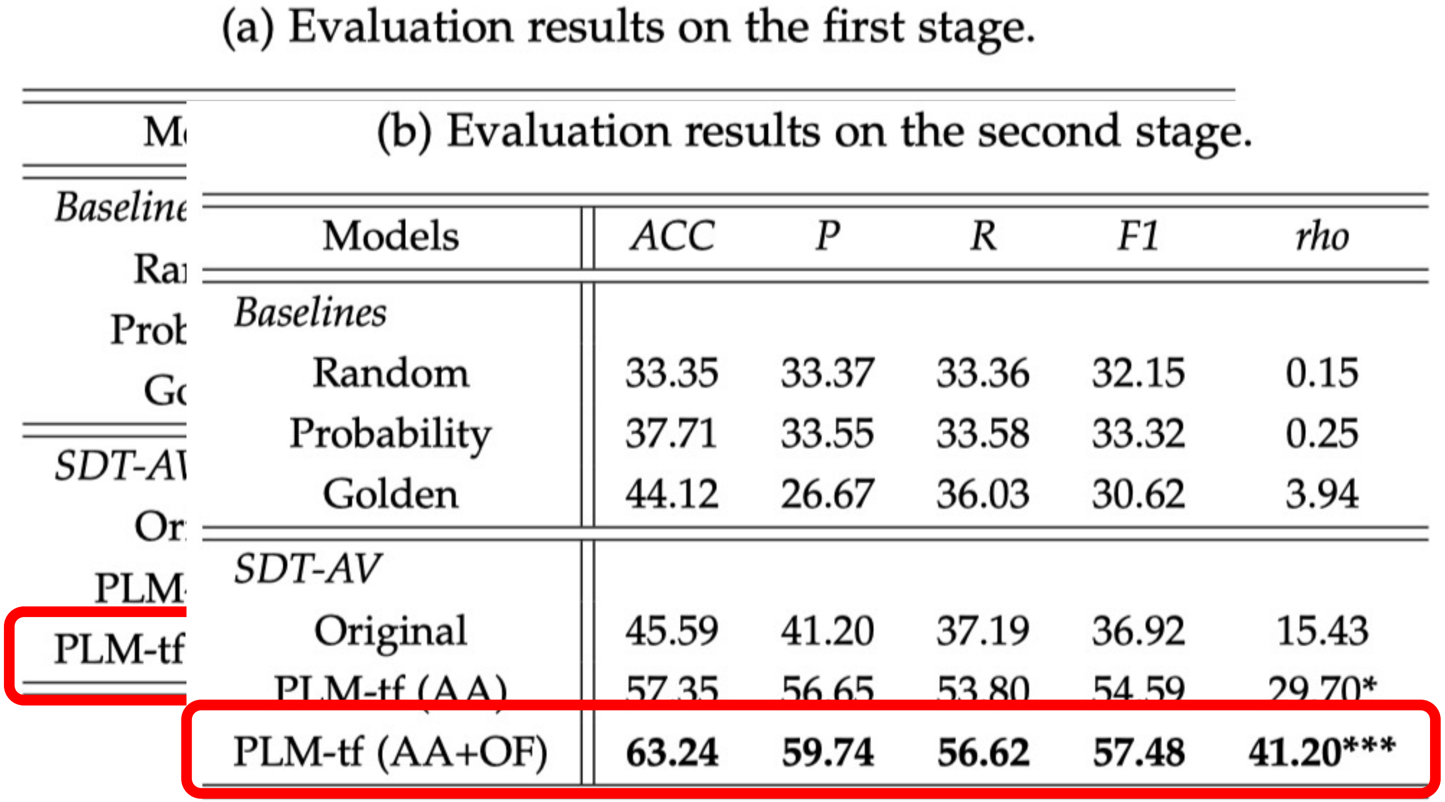


(a) Evaluation results on the first stage.

Models	*ACC*	*P*	*R*	*F1*	*rho*
Baselines					
Random	33.27	33.21	33.25	32.27	0.07
Probability	36.14	33.24	33.26	33.00	-0.68
Golden	38.24	24.47	36.51	28.79	14.91
SDT-AV					
Original	33.82	27.36	28.21	27.09	16.31
PLM-tf (AA)	51.47	50.71	**51.11**	50.30	38.75**
PLM-tf (AA+OF)	**54.41**	**50.94**	50.08	**50.37**	**38.96****

Results of the computational models

Comparison on the Outer Loop Cross-Validation of Nested-LOOCV with Baselines



(a) Evaluation results on the first stage.

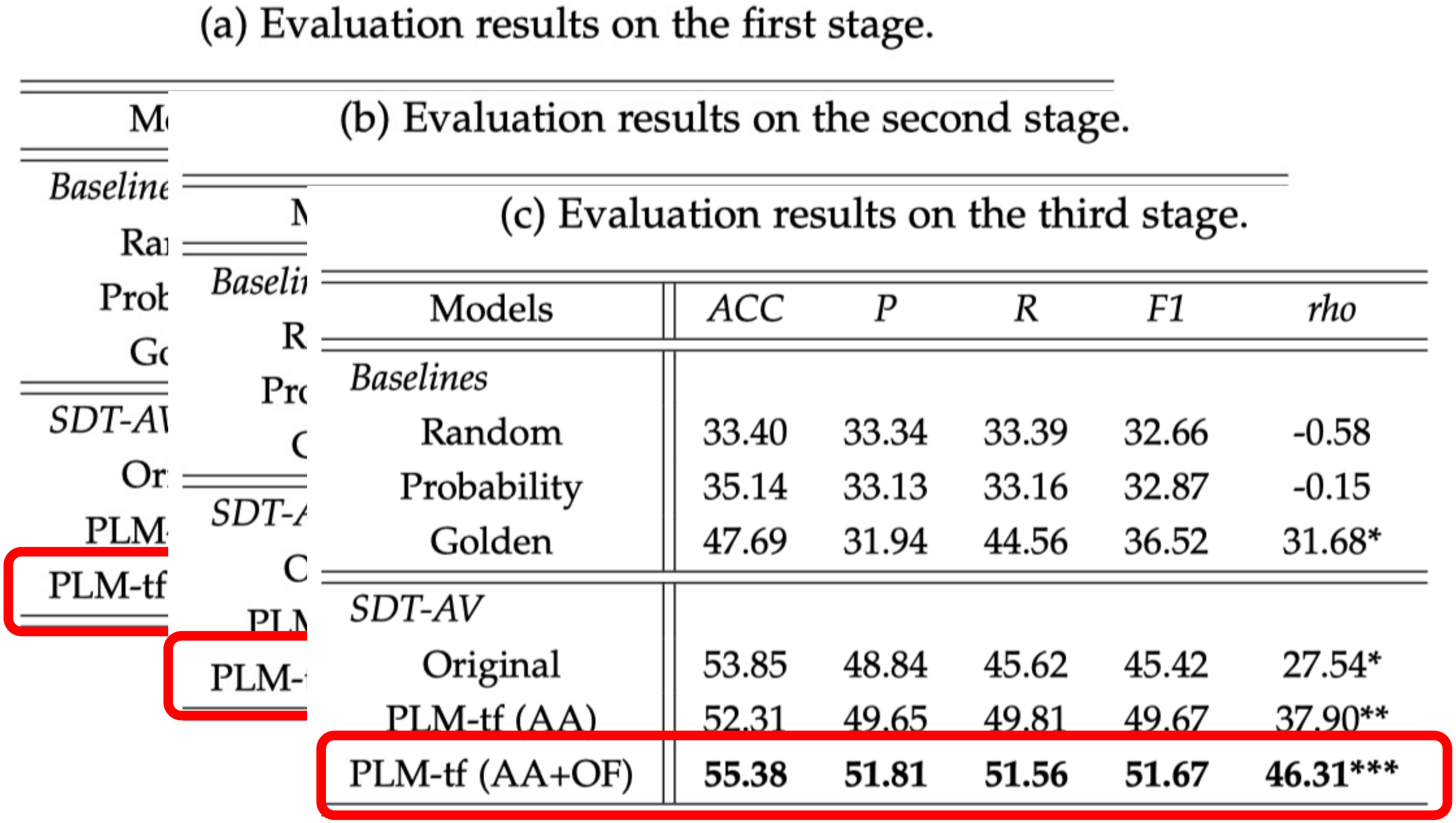
M
Baseline
Ra
Prob
Go
SDT-AV
Or
PLM-
PLM-tf

(b) Evaluation results on the second stage.

Models	*ACC*	*P*	*R*	*F1*	*rho*
Baselines					
Random	33.35	33.37	33.36	32.15	0.15
Probability	37.71	33.55	33.58	33.32	0.25
Golden	44.12	26.67	36.03	30.62	3.94
SDT-AV					
Original	45.59	41.20	37.19	36.92	15.43
PLM-tf (AA)	57.35	56.65	53.80	54.59	29.70*
PLM-tf (AA+OF)	**63.24**	**59.74**	**56.62**	**57.48**	**41.20*****

Results of the computational models

Comparison on the Outer Loop Cross-Validation of Nested-LOOCV with Baselines



(a) Evaluation results on the first stage.

M
Baseline
Ra
Prob
Go
SDT-AV
Or
PLM
PLM-tf

(b) Evaluation results on the second stage.

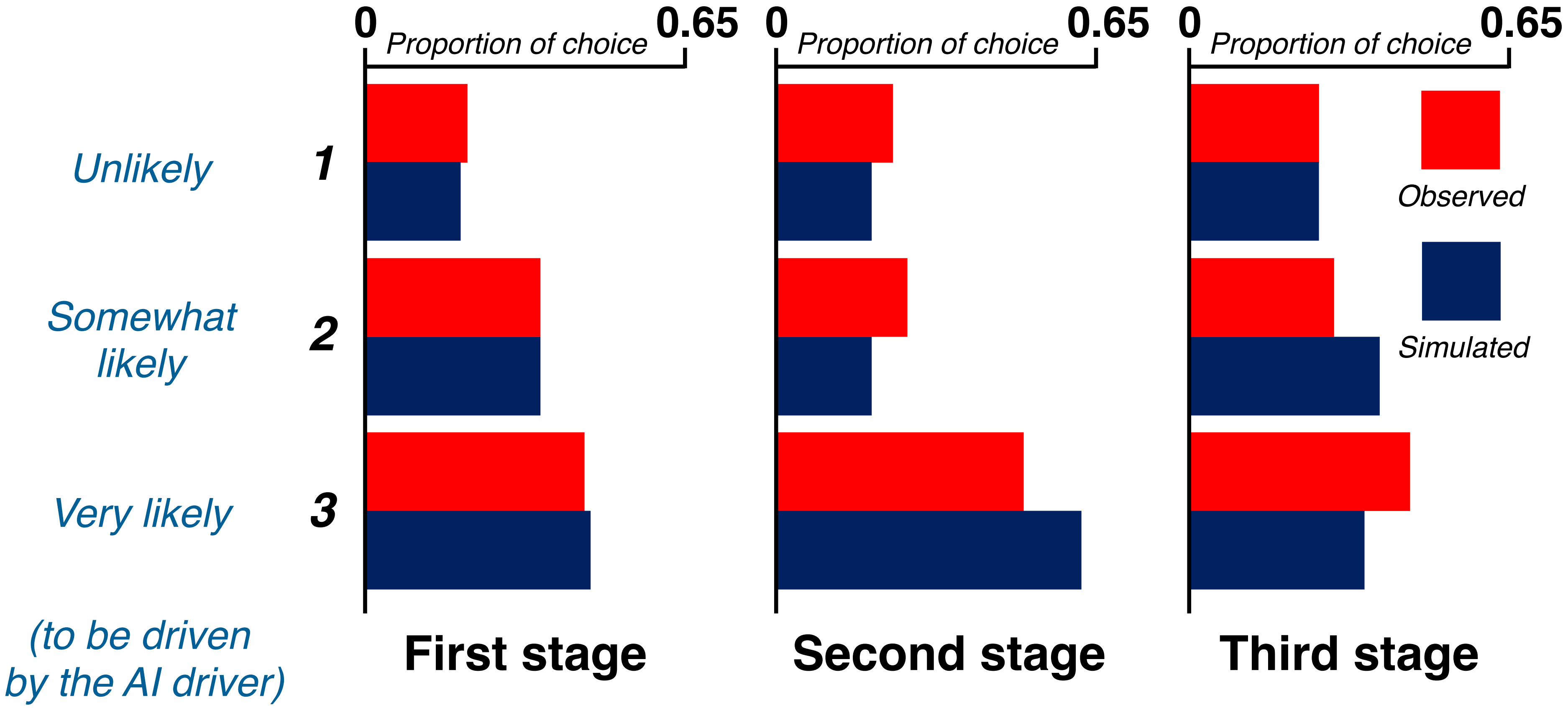
M
Baseli
R
Pro
C
SDT-
C
PLM
PLM-

(c) Evaluation results on the third stage.

Models	*ACC*	*P*	*R*	*F1*	*rho*
Baselines					
Random	33.40	33.34	33.39	32.66	-0.58
Probability	35.14	33.13	33.16	32.87	-0.15
Golden	47.69	31.94	44.56	36.52	31.68*
SDT-AV					
Original	53.85	48.84	45.62	45.42	27.54*
PLM-tf (AA)	52.31	49.65	49.81	49.67	37.90**
PLM-tf (AA+OF)	**55.38**	**51.81**	**51.56**	**51.67**	**46.31*** **

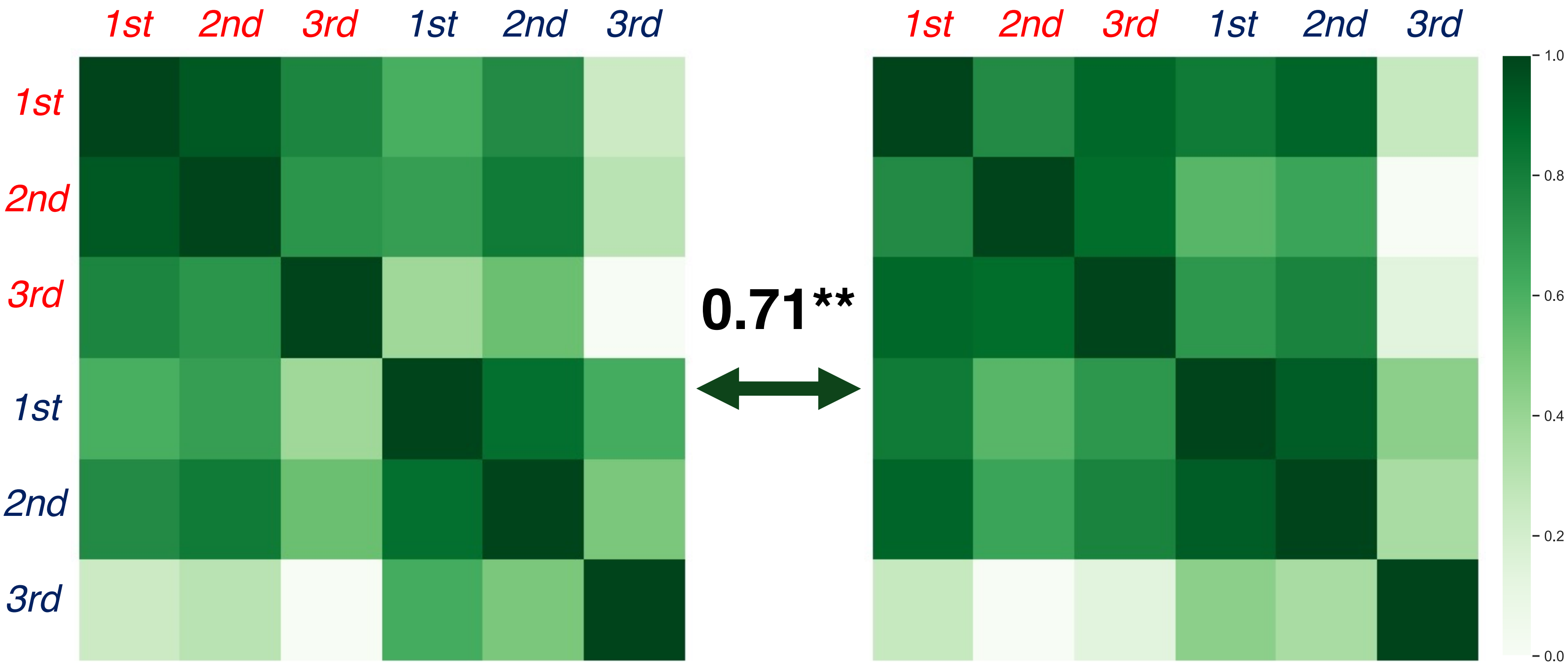
Results of the computational models

Comparison of the proportion of choices between model simulations (blue) and empirically observed choices (red)



Results of the computational models

Representational similarity between the representational similarity matrix (RSM) of empirically observed choices (left) and model simulations (right) averaged over all participants.



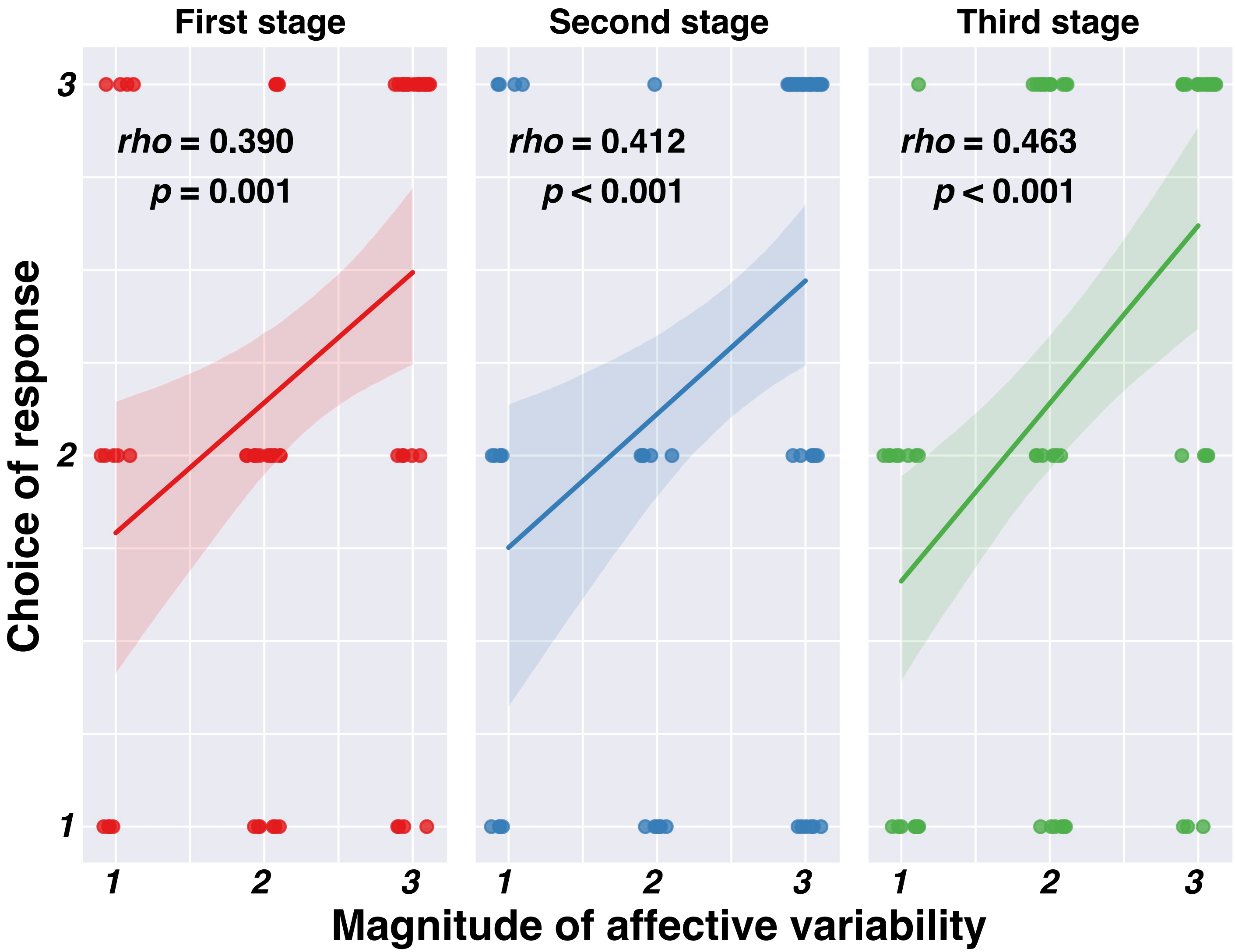
Observed RSM

Simulated RSM

Correlations between choice of response and affective variability

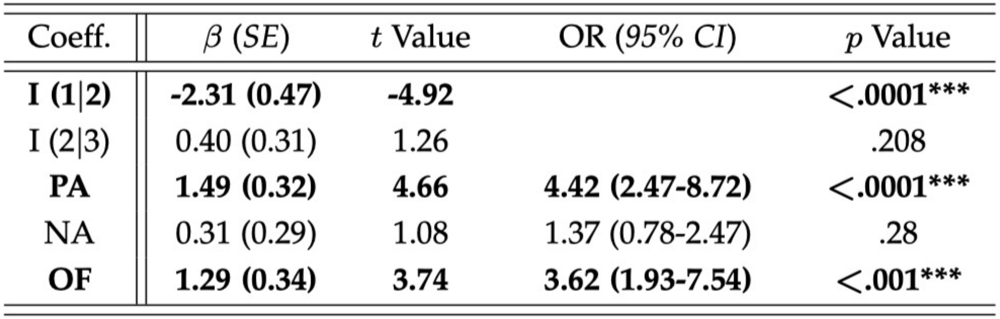
The Spearman's rank correlation score between

the gold labels and the magnitude of affective variability (AV)



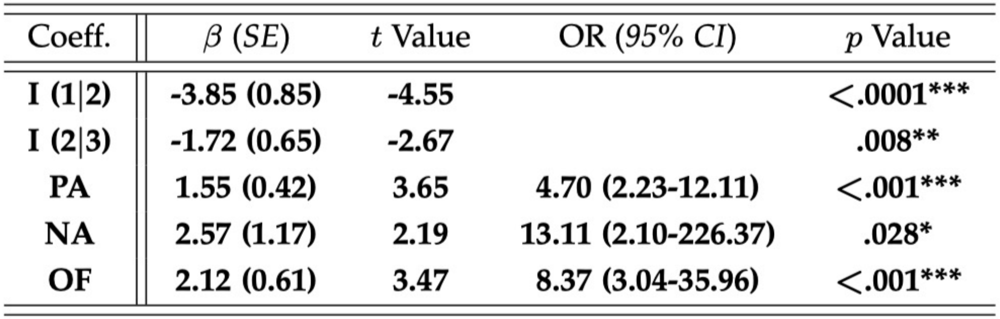
Ordinal logistic regression analysis of model simulations

(a) Results of OLR predicting simulated labels on the first stage.



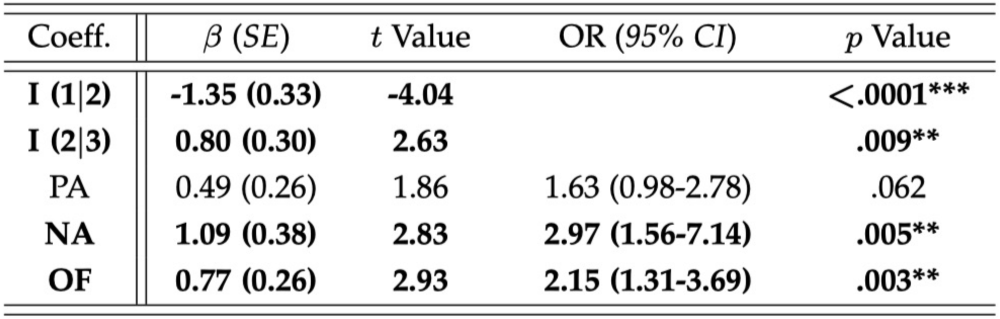
Coeff.	*β* (*SE*)	*t* Value	OR (*95% CI*)	*p* Value
I (1|2)	**-2.31 (0.47)**	**-4.92**		**<.0001*****
I (2|3)	0.40 (0.31)	1.26		.208
PA	**1.49 (0.32)**	**4.66**	**4.42 (2.47-8.72)**	**<.0001*****
NA	0.31 (0.29)	1.08	1.37 (0.78-2.47)	.28
OF	**1.29 (0.34)**	**3.74**	**3.62 (1.93-7.54)**	**<.001*****

(b) Results of OLR predicting simulated labels on the second stage.

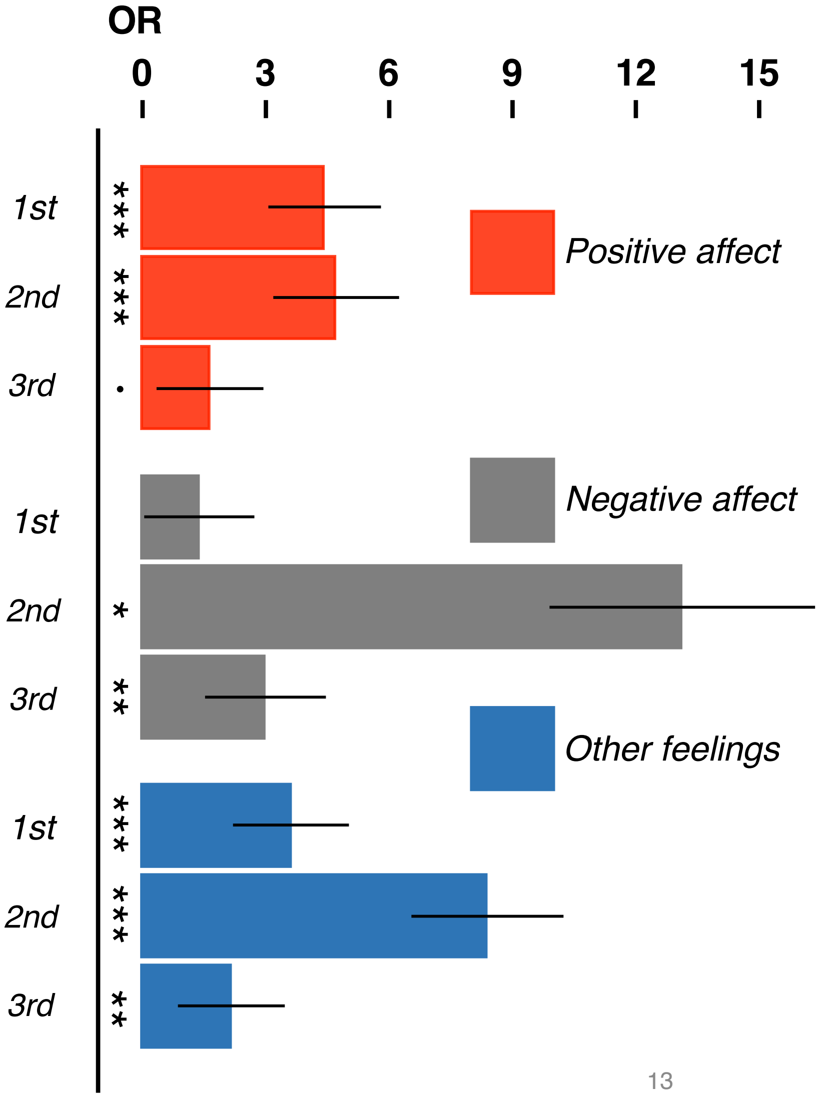


Coeff.	*β* (*SE*)	*t* Value	OR (*95% CI*)	*p* Value
I (1|2)	**-3.85 (0.85)**	**-4.55**		**<.0001*****
I (2|3)	**-1.72 (0.65)**	**-2.67**		**.008****
PA	**1.55 (0.42)**	**3.65**	**4.70 (2.23-12.11)**	**<.001*****
NA	**2.57 (1.17)**	**2.19**	**13.11 (2.10-226.37)**	**.028***
OF	**2.12 (0.61)**	**3.47**	**8.37 (3.04-35.96)**	**<.001*****

(c) Results of OLR predicting simulated labels on the third stage.



Coeff.	*β* (*SE*)	*t* Value	OR (*95% CI*)	*p* Value
I (1|2)	**-1.35 (0.33)**	**-4.04**		**<.0001*****
I (2|3)	**0.80 (0.30)**	**2.63**		**.009****
PA	0.49 (0.26)	1.86	1.63 (0.98-2.78)	.062
NA	**1.09 (0.38)**	**2.83**	**2.97 (1.56-7.14)**	**.005****
OF	**0.77 (0.26)**	**2.93**	**2.15 (1.31-3.69)**	**.003****



Discussion and conclusion

Contributions and implications

In the present study, for the first time, we examined whether the current SAE Level 4 AC could pass the modified Turing test of automated driving from the perspective of passive passengers in a real road scenario.

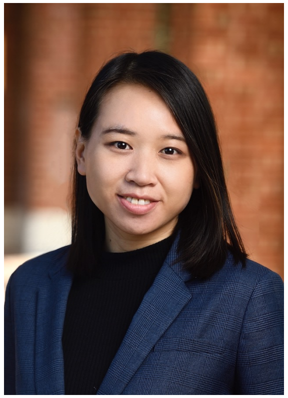
On the basis of the classical Lewin's equation, we propose a model combining SDT with AV (transformed by PLMs) to predict the passenger's choice behaviour in the Turing test. This is, to the best of our knowledge, the first computational model which provides a mechanistic understanding underlying passengers' mentalising process.

Our results shed light on the direction of future automated driving, which should improve the affective stability of passengers. Considering the fact that machines take on increasingly social roles, our suggestion may not be limited to automated driving but the whole realm of human machine interactions.

Acknowledgement & contact

Qiaoli
Jiang

Zhengming
Wu



Anqi
Liu

Haiyan
Wu

Miner
Huang

Kai
Huang

Yixuan
Ku

@ANDlab3

***A**ffective*
***N**euroscience and*
***D**ecision-making*
Lab
andlab-um.com

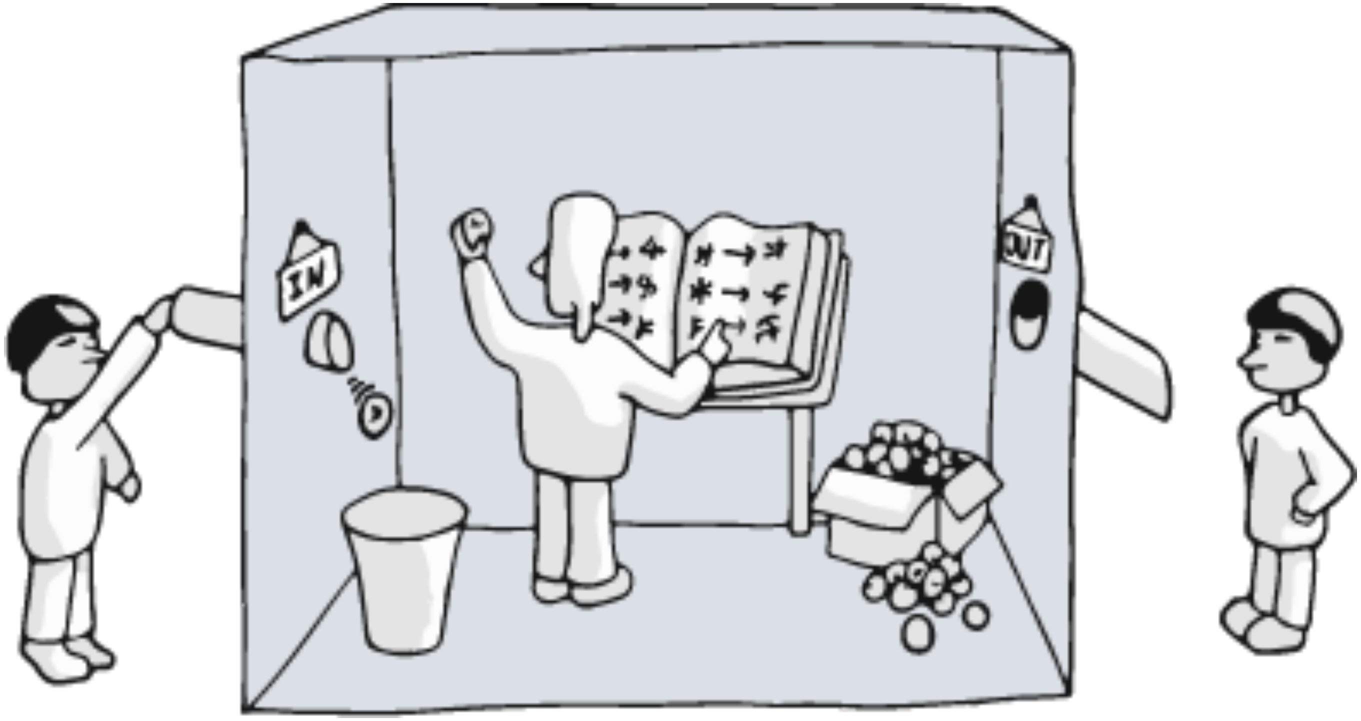
A.N.D
Lab

Presenter email: yc17319@umac.mo @lizhn7

Discussion and conclusion

Limitations and future work

While our results showed the AI driver passed the Turing tests, we will not go so far as to suggest that the AI driver “thinks” like a human driver.



Searle’s Chinese room thought experiment

(Adapted from Wikipedia)

Discussion and conclusion

Limitations and future work

While our results showed the AI driver passed the Turing tests, we will not go so far as to suggest that the AI driver “thinks” like a human driver.

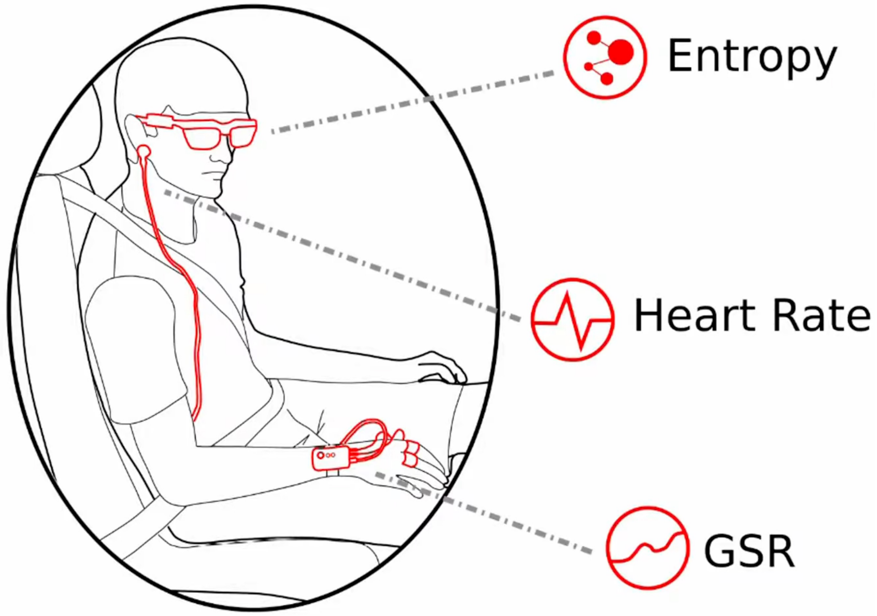
We just focused on the modified Turing test for the narrow or weak AI agent in the non-social context.

A validation test would be crucial in future work to test whether our findings will remain.

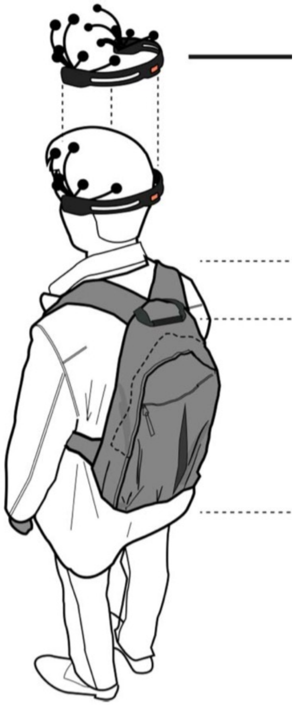
Discussion and conclusion

Limitations and future work

We only used self-reported scores to measure the emotion experiences of passengers, which limits our adventure towards the brain mechanisms supporting passengers' mentalising process in the Turing test.



(Dillen et al, 2020)



(Aspinall et al, 2013)

(Piper et al, 2014)